

Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Fundamentos de Big Data

Lección 5

Kaggle y los retos de Data Science

ÍNDICE

- | **Introducción**
- | **Objetivos**
- | **Buscamos los “missing values”**
- | **Predicciones para diferentes algoritmos**
- | **Submission en Kaggle**

INTRODUCCIÓN

En esta lección haremos una breve introducción a Kaggle y los retos de Data Science a través del Titanic Dataset.

OBJETIVOS

Al finalizar esta lección serás capaz de:

- 1 Entender los pasos a seguir para analizar el Titanic Dataset
- 2 Comprender conceptos como el escalado de los datos, “missing values”, etc.
- 3 Enviar resultados a Kaggle para valorar tus predicciones

Buscamos los “missing values”

```
In [58]: df.isnull().sum()
```

```
Out[58]: Survived      0  
Pclass      0  
Name        0  
Sex         0  
Age        177  
SibSp       0  
Parch       0  
Ticket      0  
Fare        0  
Cabin      687  
Embarked    2  
dtype: int64
```

Predicciones para diferentes algoritmos

Pruebo posibles algoritmos

```
In [82]: # KNeighborsClassifier
clf = KNeighborsClassifier()
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
acc_KN = accuracy_score(y_test, y_pred)
acc_KN
```

Out[82]: 0.8212290502793296

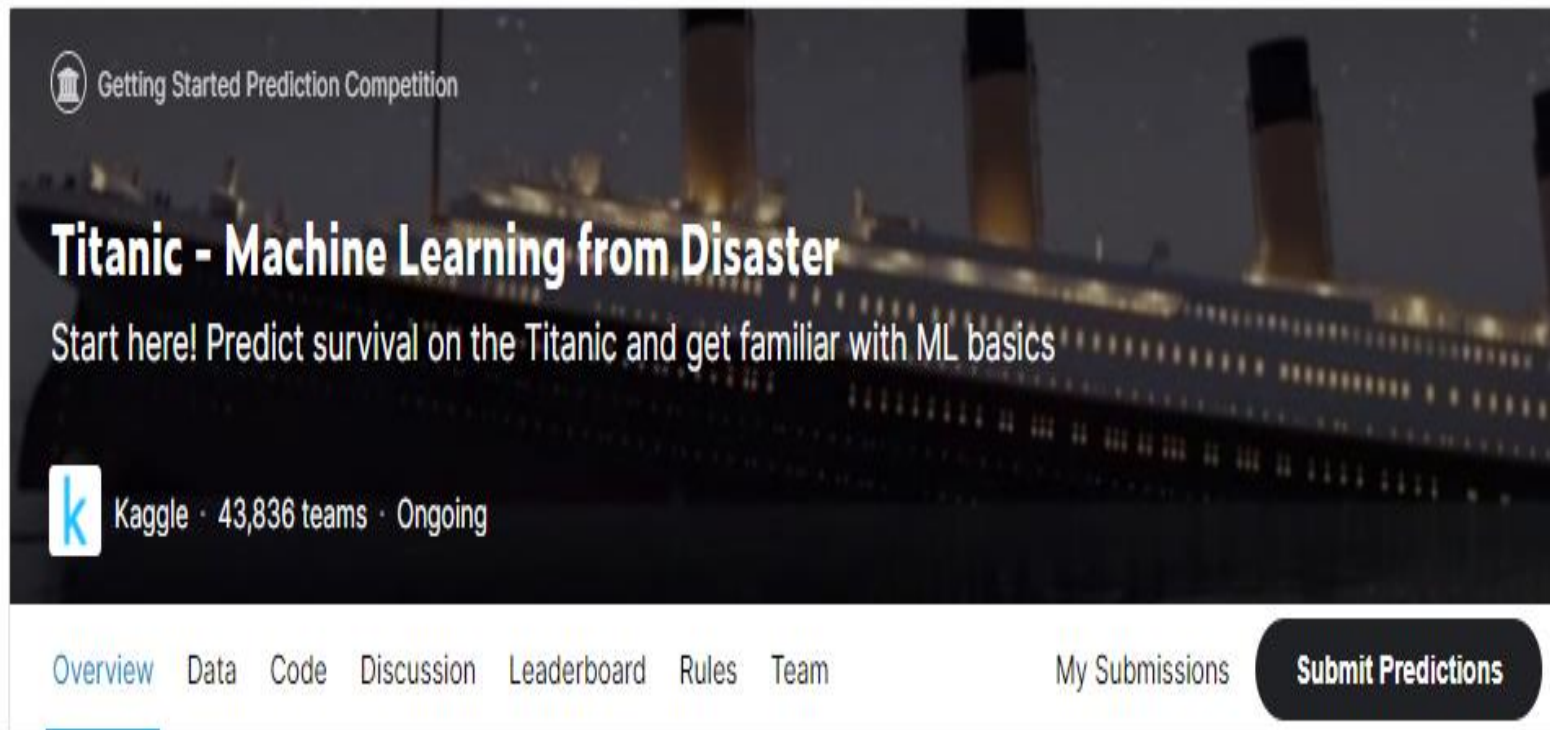
```
In [83]: # DecisionTreeClassifier
clf = DecisionTreeClassifier()
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
acc_DT = accuracy_score(y_test, y_pred)
acc_DT
```

Out[83]: 0.776536312849162

```
In [84]: # RandomForestClassifier
clf = RandomForestClassifier()
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)
acc_RF = accuracy_score(y_test, y_pred)
acc_RF
```

Out[84]: 0.8212290502793296

Submission en Kaggle




The screenshot shows the Kaggle competition page for 'Titanic - Machine Learning from Disaster'. The background is a dark image of the Titanic ship at night. At the top left, there is a 'Getting Started Prediction Competition' icon. The main title 'Titanic - Machine Learning from Disaster' is in large white font, followed by the subtitle 'Start here! Predict survival on the Titanic and get familiar with ML basics'. Below this, the Kaggle logo is shown next to the text 'Kaggle · 43,836 teams · Ongoing'. At the bottom, there is a navigation bar with links: 'Overview' (underlined), 'Data', 'Code', 'Discussion', 'Leaderboard', 'Rules', and 'Team'. To the right of these links are 'My Submissions' and a large black button labeled 'Submit Predictions'.

Getting Started Prediction Competition

Titanic - Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

 Kaggle · 43,836 teams · Ongoing

[Overview](#) [Data](#) [Code](#) [Discussion](#) [Leaderboard](#) [Rules](#) [Team](#) [My Submissions](#) [Submit Predictions](#)

MUCHAS GRACIAS POR SU ATENCIÓN



jmpena@grupomainjobs.com



José Manuel Peña

<https://www.linkedin.com/in/jos%C3%A9-manuel-pe%C3%B1a-castro-7566b349/>



twitter.com/eiposgrados



facebook.com/eiposgrados



instagram.com/eiposgrados